Practical Operational Risk

Rosella Giacometti[1]

Associated Professor in Mathematical Finance
Department of Mathematics, Statistics, Computer Science and Applications "L.Mascheroni"
Faculty of Economics and Business Administration
Via dei Caniana 2, 24122 Bergamo, Italy
E-mail: rosella.giacometti@unibg.it

Svetlozar Rachev[2]

Chair-Professor, Chair of Econometrics, Statistics
and Mathematical Finance
School of Economics and Business Engineering
University of Karlsruhe
Postfach 6980, 76128 Karlsruhe, Germany
and
Department of Statistics and Applied Probability
University of California, Santa Barbara
CA 93106-3110, USA
E-mail: rachev@statistik.uni-karlsruhe.de

Anna Chernobai

Assistant Professor of Finance
Department of Finance
M.J. Whitman School of Management
Syracuse University
721 University Ave, Room 609
Syracuse, NY 13244-2450,USA
E-mail: annac@syr.edu

Marida Bertocchi

Full Professor in Mathematical Finance
Department of Mathematics, Statistics, Computer Science and Applications "L.Mascheroni"
Faculty of Economics and Business Administration
Via dei Caniana 2, 24122 Bergamo, Italy
E-mail: marida.bertocchi@unibg.it

Giorgio Consigli

Associated Professor in Mathematical Finance
Department of Mathematics, Statistics, Computer Science and Applications "L.Mascheroni"
Faculty of Economics and Business Administration
Via dei Caniana 2, 24122 Bergamo, Italy
E-mail: giorgio.consigli@unibg.it

---

[1] Rosella Giacometti's research was supported by grant from the Department of Mathematic, Statistica, Computer Science and Application MIUR 60% 2006 "Misurazione e gestione dei rischi operativi " 60GIAC07.

[2] Svetlozar Rachev's research was supportes by grants from Division of Mathematical, Life and Physical Sciences, College of Letters and Science, University of California, Santa Barbara, and the Deutschen Forschungsgemeinschaft.

Practical Operational Risk

Abstract

Operational losses are generally observed at specific points in time and vary from moderate to possibly very large amounts. Both variables – the time of the event and the amplitude of the associated loss – are random variables whose distributions must be estimated.

The concept of **compound Poisson process** provides an accurate analytical framework to address the modelling problem. The time of the event is captured by the intensity of the Poisson process and losses by an appropriate distribution. The aggregated operational losses are in this setup considered to follow a random process $\{L(t)\}_{t\geq 0}$ with

$$L_t = \sum_{k=0}^{N_t} X_k, \quad X_k \propto F_\theta, N_t \propto Poi(\lambda).$$

The operational loss distribution is thus jointly determined by the average number of losses per unit of time – the intensity λ of the Poisson process N_t, the counting process with integer variables – and by the loss magnitudes X_k – in monetary terms – observed over time. X_k are in general assumed to belong to a family F_θ of parametric continuous distributions.

In this paper, we analyse the class of parametric distributions which better *fit* the observed empirical loss data classified by business lines. Particular attention is devoted to the fitting of the tail distribution of the losses.

1. INTRODUCTION

In the course of the last few decades, the financial industry has been characterised by accelerated globalisation, deregulation, and a boost of new products, instruments, and services. An inevitable outcome has been an elevated exposure of the financial institutions to various sources of risk. A large proportion of these financial risks are attributed neither to market nor to credit risks, and it is known as operational risk. Examples include losses due to unauthorised trading, fraud, human errors on the Orange County (USA, 1994), Barings (Singapore, 1995), Daiwa (Japan, 1995), among many others. It is often the risk connected to large magnitude losses that rarely occur (i.e., low frequency, high-severity losses).

Operational risk is largely a firm-specific non-systematic risk: according to the Bank of International Settlements (1998), *"Unlike market and perhaps credit risk, the [operational] risk factors are largely internal to the bank."*

Earlier references on operational risk defined operational risk as *"other risks"*, or *"any risk not categorizes as market and credit risk"*, and *"the risk of loss arising from various types of human or technical errors"* (BIS (1998)). Other definitions include: risk *"arising from human and technical errors and accidents"* (Jorion (2000)) *"a measure of the link between a firm's business activities and the variation in its business results"* (King (2001)) and *"the risk associated with operating a business"* (Crouhy et al. (2001)).
The Basel II Capital Accord of 2001 (BIS (2001a,b), (2003), (2004)) defined operational risk as "the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events". This definition includes legal risk, but excludes strategic and reputational risk.

Current estimates suggest that the allocation of total financial risk of a bank is roughly 60% of their regulatory capital to credit risks, 15% to market risks, and 25% to operational risk (Jorion (2000)). Cruz (2002) suggests different figures, 50%, 15%, 35% respectively. Under the 2001 Basel II Capital Accord (BIS (2001a)) each bank is required to adopt a methodology to determine the operational risk capital charge to account for unexpected losses by the end of 2006. BIS suggested the following methodologies: Basic Indicator Approach, Standardized Approach, and three Advanced Measurement Approaches (Internal Measurement Approach, Loss Distribution Approach, and Scorecard Approach). The choice of the method depends on bank's size and risk exposure and the ability to meet required criteria. The Loss Distribution Approach (hereforth, LDA) is the most accurate from the statistical point of view as it utilises the exact distribution of the losses – both frequency and severity – based on each individual bank's internal loss data. Under the LDA, the operational risk capital charge is determined by the value-at-risk measure (hereforth, VAR), summed across all "business line/event type" combinations. The capital charge is hence dependent on the nature of the frequency and severity distributions. In particular, under the VAR measure, the tails of these distributions become the key determinants of the amount of the capital charge.

The Basel II Capital Accord classifies operational risk into 7 event type groups, 8 business lines, and identifies 6 operational loss types, see BIS (2003). Detailed descriptions of business lines mapping, loss types, and event types are presented in Table 1, Table 2, and Table 3, respectively.

Business Unit	*Business Line*
Investment Banking	Corporate Finance Trading and Sales
Banking	Retail Banking Commercial Banking Payment and Settlement Agency Services
Others	Asset Management Retail Brokerage

Table 1: Business line mapping according to the Basel II Capital Accord.

Loss Type	*Contents*
Write-downs	Direct reduction in value of assets due to theft, fraud, unauthorized activity or market and credit losses arising as a result of operational events
Loss of recourse	Payments or disbursements made to incorrect parties and not recovered
Restitution	Payments to clients of principal and/or interest by way of restitution, or the cost of any other form of compensation paid to clients
Legal Liability	Judgments, settlements and other legal costs
Regulatory and Compliance	Taxation penalties, fines, or the direct cost of any other penalties, such as license revocations
Loss of or damage to assets	Direct reduction in value of physical assets, including certificates, due to some kind of accident (e.g., neglect, accidents, fires, earthquakes)

Table 2. Loss types and definitions according to the Basel II Capital Accord.

Event Type	Definition and Categories
Internal Fraud	Acts intended to defraud, misappropriate property or circumvent regulations, the law or company policy, which involves at least one internal party. Categories: unauthorized activity and theft and fraud.
External Fraud	Acts of a type intended to defraud, misappropriate property or circumvent the law, by a third party. Categories: (1) theft and fraud and (2) systems security.
Employment Practices and Workplace Safety	Acts inconsistent with employment, health or safety laws or agreements, from payment of personal injury claims, or from diversity/discrimination events. Categories: (1) employee relations, (2) safe environment, and (3) diversity and discrimination.
Clients, Products, and Business Practices	Unintentional or negligent failure to meet a professional obligation to specific clients (including fiduciary and suitability requirements), or from the nature or design of a product. Categories: (1) suitability, disclosure and fiduciary, (2) improper business or market practices, (3) product flaws, (4) selection, sponsorship and exposure, and (5) advisory activities.
Damage to Physical Assets	Loss or damage to physical assets from natural disaster or other events. Categories: Disasters and other events.

Business Disruption and System Failures	Disruption of business or system failures. Categories: systems.
Execution, Delivery, and Process Management	Failed transaction processing or process management, from relations with trade counterparties and vendors. Categories: (1) transaction capture, execution and maintenance, (2) monitoring and reporting, (3) customer intake and documentation, (4) customer/client account management, (5) trade counterparties, and (6) vendors and suppliers.

Table 3. Event types and descriptions according to the Basel II Capital Accord.

In this paper we analyze the operational risk of a Bank using both the internal data supplied by the Bank and integrating, when necessary, with an external database. We focus on the following aspects of the operational risk:

- We apply the canonical statistical analysis (descriptive statistics, alternative parametric fitting, goodness-of-fitting tests) to internal and external loss data, separately and then with a specifically tailored integration procedure,
- We explore the implications of a loss estimation approach based on mixed distributions with an extreme value distribution (EVD, such as the GPD) explicitly modelling only the extreme losses.

The paper is structured as follows. Section 1. is devoted to methodologies for assessing operational risk. In Section 2, we present our case study, describe the procedure adopted to integrate internal, and external losses and present the statistical evidence for the determination of the severity conditional distributions and the frequency distribution for each one of the three BL's included in our study. We extend the analysis in order to show how it is possible to improve the fitting in the far tail introducing a mixed distribution.

2. A Methodology For Assessing Operational Risk: Modification of The LDA Approach

Operational risk possesses unique characteristics that distinguish it from other sources of financial risk. The nature of operational risk is very different from that of market risk and credit risk, see King (2001). In fact, operational losses share many similarities with insurance claims, suggesting that most actuarial models can be a natural choice of the model for operational risk, and models well developed by the insurance industry can be almost exactly applied to operational risk, see Bening and Korolev (2002), Chernobai et al. (2005a), Grandell (1991), Panjer and Wilmott (1992), Thorin and Wikstad (1977).

The LDA approach is computationally intensive but has the advantage of taking into account the frequency and severity distributions. The two distributions are first estimated individually, then the aggregate distribution is computed using the compound Poisson process, and finally the appropriate risk measures are added across all "business line/event type" combinations This methodology implicitly assumes that "business line/event type" combinations are perfectly correlated.

A number of problems arise when one tries to deal with operational risk in practice. First, banks collect data losses only for losses above a certain threshold that may vary from bank to bank. It means that if one fits the data without considering the missing data, results would be biased. Second, most banks began collecting operational risk data only from 2001 and given the relatively low frequency of operational risk events, one may face the problem of an insufficient set of data.

The third problem deals with the need to supplement internal data with external data in order to improve the accuracy of the statistical measurements (Baud, Frachot and Roncalli (2002)). Generally, first generation external database only record the highest losses, i.e., the losses that are publicly released, while consortium-based data are anonymised data collected by a consortium of banks, see for example the ORX project (Peemoller (2002)). It means that pooling together internal and external data without adopting special statistical treatment may result in capital estimates that are over-stated. One characteristic common to both internal and external databases is the presence of a lower collection threshold which is generally different in the two databases. Thus a right methodology is needed in order to have data comparable.

The fourth problem is concerned with the reality that operational losses may possess some dependence structure. For example, a failure in a bank's computer system may interrupt its important financial transfers. However, in general the correlation is spurious and it is not a dominant feature, see Roher (2002). Therefore, in what follows the hypothesis of independent and identically distributed (i.i.d) data is considered.

Finally, the recording date of an operational loss can be related to one of the three dates: the date when the loss took place (date of event occurrence), the date on which it was revealed that the event has taken place, and the date on which the loss amount was recorded. In our bank's internal database, we are using the date corresponding to the date on which the loss took place.

2.1 Loss Frequency Process

One of the difficulties that arise with modeling operational losses has to do with the irregular nature of the event arrival process. Operational losses occur at irregular time intervals suggesting a process of a discrete nature. This makes it similar to the reduced-form models for credit risk, in which the frequency of default (i.e., failure to meet a credit agreement) is of non-trivial concern.

One can adopt one of the two modelling methodologies: modeling the counting process, i.e., the distribution of the number events in a fixed time interval, or modeling the inter-arrival times' distribution. We adopt the first approach.

It is reasonable to assume that, in most situations, operational risk-related events arrive independently from each other. A common model to characterize such a process is a Poisson process[3] (Bening and Korolev (2002), Grandell (1997), Wilmott (1990)). In the simplest

---

[3] *We recall the definition: of a Poisson process*. A stochastic process N_t, $t\geq 0$, is called Poisson process if it satisfies the following properties:
1. N_t has independent increments, i.e., for any natural n, any $t_0, t_1, \dots, t_n$, such that $0\leq t_1\leq t_2\leq\dots<\infty$, the random variables $N_{t1}-N_{t0}, N_{t2}-N_{t1}, \dots, N_{tn}-N_{tn-1}$ are independent; 2. N_t is homogeneous, i.e., for any s_0, t_0 and $h>0$, the random variables $N_{t+h}-N_t$ and $N_{s+h}-N_s$ are identically distributed;
3. $N_0 = 0$;
4. the number of jumps in an interval t is Poisson distributed with mean λt, $\lambda>0$, i.e., for all t, s > 0,

$$P(N_{t+s}-N_s=n)=\frac{(\lambda t)^n e^{-\lambda t}}{n!}, \quad n=0,1,\dots$$

scenario, the mean number of events per unit of time is constant in time. In practice, however, it is plausible to expect that the mean number of events in a given time interval does not remain constant but behaves in a more random fashion or even evolves and changes with time. For the counting process we are testing three possible discrete distributions, the Poisson distribution, Poisson Process with a lognormal-cdf like cumulative intensity and Poisson Process with a logweibull-cdf like cumulative intensity.

If λ (i.e., the intensity rate) is a constant, we have a homogeneous Poisson process (HPP) that has a cumulative intensity λt. The mean of a homogeneous Poisson distribution equals the variance. When λ is not constant, we have a non-homogeneous Poisson process (NHPP).

In non-homogeneous Poisson processes, λ is believed to evolve with time in a fashion that can be expressed by a mathematical function, λ(t). For example, a possible cyclical component in the time series of the number of loss events may be captured by a sinusoidal rate function, an upward-sloping tendency may be captured by a quadratic function, and so on. Moreover, deviations from an assumed (or fitted) deterministic model may be further captured by a random stochastic process, such as Brownian motion.

In our case we model the cumulative number of losses by the following Non-Homogenuos Poisson Process with cumulative intensity:

Type I lognormal-cdf like

$$\Lambda(t) = a + \frac{b \exp\left(\frac{-\log^2 (t-d)}{2c^2} \right)}{\sqrt{2\pi c}}$$

Type II logweibull-cdf like

$$\Lambda(t) = a - b \exp(-c \log^d (t)).$$

2.2 Loss Severity Process

A variety of loss distributions can be used to model operational loss magnitudes: Lognormal, Gamma, Weibull, Logweibull, Generalised Pareto, Burr, Symmetric α-Stable (SymStable), and log α-stable. Heavy-tailed loss distributions such as Logweibull, Generalised Pareto, SymStable, and log α-stable (Rachev (2000), Rachev and Mittnik (2000), Rachev et al. (1998), Samorodnitsky (1994)) are expected to provide a superior fit (Chernobai et al. (2005b,c)).

We can estimate the Maximum Likelihood parameters of all these distributions and assess the choice of the more appropriate severity distributions with in-sample goodness-of-fit tests. We call this approach the naïve approach because it forgets that operational risk data are incomplete data as banks record losses only above a certain threshold. As previously stated, operational loss data are subject to minimum collection threshold which is a fixed pre-determined amount. For the data set used in this paper, the threshold is set at approximately Euro 500 for the internal database and approximately Euro 5,000 for the external database. We can characterise operational risk data as “left-truncated” data since neither the number (i.e., the frequency) nor the values (i.e., the severity) of such observations have been recorded. The incomplete data refers to the recorded observations all of which fall above a positive threshold of a specific amount.

Empirical studies have shown that ignoring the incompleteness of the data may result in severe under-estimation of the operational risk capital charge (Moscadelli et al. (2005)), creating a so-called reporting bias.

In this paper, the incompleteness of data is explicitly taken into account in the estimation of the severity distribution. We call it the conditional approach. The severity is indeed estimated "conditionally" on the fact that the observed data are now recognised as actually truncated data set and no longer a complete data set. Under the reasoning, the truncated loss distribution is fitted to the severity data, with the density expressed as follows:

$$f_\theta^c(x) = f_\theta(x \mid x \geq H) = \begin{cases} \dfrac{f_\theta(x)}{1-F_\theta(H)} & x \geq H \\ 0 & x < H \end{cases}$$

where H is the known threshold, θ is the unknown parameter set, $f_\theta(x)$ is the probability density function and $F_\theta(x)$ is the cumulative density function (in the following referred respectively as pdf and cdf).
The frequency parameter is adjusted according to the estimated fraction of the data over the threshold, which is obtained using the parameters of the fitted conditional severity distribution In general, such fraction of data may be estimated on the basis of the pertinent value of the severity distribution. Indeed, under the true severity distribution, each data point (or, rather, each range of data since we are dealing with continuous distributions) would have a probability of falling under the threshold H equal to the distribution function computed at *H*, that is *F(H)*, and *1- F(H)* is the probability of falling over the threshold *H*.

The unknown parameter set can be estimated in two ways:

i. Using the Maximum Likelihood estimation procedure, the parameter set is estimated by directly maximizing the constrained log-likelihood function:

$$\hat{\theta}^c_{MLE} = \arg\max_\theta \log \prod_{j=1}^{n} \frac{f_\theta(x_j)}{1-F_\theta(H)}.$$

ii. Using the Expectation-Maximization algorithm, see Dempster et al. (1977), McLachlan and Krishnan (1997), Meng and van Dyk (1997).

The intensity rate λ(t) can be scaled up using the following transformation:

$$\hat{\lambda}(t) = \frac{\hat{\lambda}(t)^{naivè}}{1-F_{\hat{\theta}^c_{MLE}}(H)}.$$

In the following we will estimate the parameter of the conditional distributions and perform the goodness-of-fit tests after necessary adjustments for the incomplete data (Chernobai et al. (2005d)). The goodness-of-fit tests are of two different type: supremum type (Kolmogorov-Smirnov, Kuiper, Anderson-Darling, max Anderson-Darling max up) and quadratic type (Anderson-Darling integral, Anderson-Darling integral up, Cramer von Mises).

2.3. EVT and Mixed Severity Distributions

Related literature points out that the tail and the body of the loss distribution do not conform to the same law and need to be modelled separately. Three approaches can be used to deal with this issue:

i. Use *Extreme Value Theory* (EVT) (Embrechts et al. (1997), Resnick (1987)) to model extreme losses that lie beyond a predetermined high threshold. Such model is referred to as the Peak Over Threshold model. Extreme losses are then assumed to follow a Generalized Pareto Distribution. The distribution is usually expressed in terms of the cumulative distribution function:

$$F(x) = \begin{cases} 1 - \left(1 + \xi \dfrac{x-\mu}{\beta}\right)^{-1/\xi} & if\ \xi \neq 0 \\ 1 - e^{-\frac{x-\mu}{\beta}} & if\ \xi = 0 \end{cases},$$

where

$$\begin{array}{ll} x \geq \mu & if\ \xi \geq 0 \\ \mu \leq x \leq \mu - \beta/\xi & if\ \xi < 0 \end{array}.$$

Empirical studies that use EVT to model extreme operational losses generally report the tail parameter $\xi>1$ indicating a very heavy right tail of the loss distribution.

ii. *Mixture distributions* dictate that the tail and the body of the distribution follow two separate laws. A mixture distribution is a weighted average of two distributions. The unknown parameters are the two parameter sets from the two separate distributions and the corresponding weights. Empirical studies have not applied such models to operational loss data explained by short historical samples. When too many parameters are to be estimated, the result is a model with low validity.

iii. *Robust models* (Huber (2004), Knez and Ready (1997), Martin and Simin (2003), Rousseeuw and Leroy (2003), and Chernobai and Rachev (2006)) dictate that extreme events are outliers and take them out of the data samples. Comparison of two models – the classical model in which all data are taken in the analysis and the robust model in which data are truncated from above – can reveal the economic impact of the highest operational losses. Limited empirical studies in this area suggest that 5% of outlying extreme events account for over 50% of the aggregate capital charge (Chernobai and Rachev (2006)).

The GPD belongs to the class of extreme value distributions (EVD) and has gained popularity as the most natural choice of distribution to fit extreme events far in the tail: in market risk and credit risk analysis the GPD has been adopted as reference parametric distribution in several applications domains. Its role in operational risk analysis is however even more natural.

Two results provide the theoretical grounds for the adoption of an EVT approach. The first result, known as the Fisher-Tippett theorem (1928) clarifies under which conditions we can expect that a limit distribution belongs to the class of EVD. The second result, credited to Balkema and de Haan (1974) and Pickands (1975), motivates the adoption of GPD as reference distribution in EVT applications. The Fisher-Tippett theorem can be regarded as the central limit theorem for the class of EVD. In this set-up, the use of GPD is then justified by the fact that if, under the same assumptions, the distribution of the *excesses*

$F_u(x) = P[X - u \le x \mid X > u], x > 0$ beyond a certain threshold u is such that an EVD exists, then it can be proven that this limit distribution is the two-parameter GPD denoted by $G_{\xi,\beta(u)}(x)$.

Unbiased estimates of the two parameters of the GPD require a sufficient sample set for the estimation procedure: this is the crucial trade-off problem of any EVT applications. The threshold needs to be sufficiently far in the tail to apply the above mentioned EVT results, while the sample size needs to be sufficiently large to allow a correct formulation and solution of the estimation problem.

In this paper, we discuss the applicability of EVT to our dataset, the procedure to determine *u* and then we apply the first two presented approaches, estimating explicitly the *excesses over a threshold u* and then introducing a mixed distribution.
Let *u* be the selected threshold for the support partition, we then define the mixed distribution as :

$$h(x) = \pi \frac{f(x)}{F(x)} 1_{x \le u} + (1-\pi) g(x) 1_{x>u}$$

where π and $1\text{-}\pi$ can be seen respectively as the importance of the body and the tail and g(x) represent the probability distribution function of the GDP distribution.

2.4 The compound Poisson Process

Under the LDA approach the frequency and severity distributions are estimated separately since they are assumed to be independent, and afterwards the aggregate loss distribution is computed by the compound Poisson process.

The concept of **compound Poisson process** provides an accurate analytical framework to address the modelling problem. The time of the event is captured by the intensity of the Poisson process and the loss by an appropriate state distribution. Then the aggregated operational losses are considered to follow a random process $\{L(t)\}_{t \ge 0}$ with

$$L_t = \sum_{k=0}^{N_t} X_k, \quad X_k \propto F_\theta, N_t \propto Poi(\lambda).$$

The operational loss distribution is thus jointly determined by the average number of losses per unit of time – the intensity λ of the Poisson process N_t, the counting process with integer variables – and by the loss magnitudes X_k – in monetary terms – observed over time. X_k are in general assumed to belong to a family F_θ of parametric continuous distributions. Finally, via Monte Carlo simulation it is possible to compute the distribution of the aggregated losses and compute expected aggregate loss (EL), unexpected aggregate loss (UL), and Value-at-Risk (VaR).

3. EMPIRICAL ANALYSIS

In this section we apply the methodology to the data provided by a large European bank. The name of the bank is not disclosed to preserve its confidentiality. In the first part of this empirical study, we will present the description of internal and external data. We will then apply a methodology to aggregate internal and external in presence of different thresholds and will estimate the frequency and the severity distributions for three business lines. Finally, we will discuss applications of EVT and mixture distributions.

3.1 INTERNAL, EXTERNAL DATA AND INTEGRATED DATA

The data set selected for the study consists of a total of slightly under 2700 observations. It is classified into three business lines: Retail Banking (business line 3 or BL3), Commercial Banking (BL4), and Retail Brokerage (BL8). The data are further classified into seven event types (ET) as presented in Table 3.[4] Table 4 describes the internal data.
BL3 accounts for 77.7% of the data of the internal database, BL4 for 7.82%, and BL8 for 8.18%.

Internal data

	BL3	BL4	BL8
ET1	0.68%	0.01%	0.01%
ET2	10.75%	0.19%	0.00%
ET3	2.19%	0.00%	0.08%
ET4	2.85%	0.57%	0.43%
ET5	1.11%	0.10%	0.00%
ET6	0.21%	0.00%	0.08%
ET7	4.25%	1.35%	1.72%
	77.70%	7.82%	8.18%

Table 4: Description of Internal Data by business lines (BL3,BL4, and BL8) and event types (ET1 to ET7)

External data are shared within the financial community and provide a reliable data source for those intermediaries that for various reasons have no or little internal loss data. External databases can, to a certain extent, be regarded as a benchmark or loss collector for the average financial intermediary.

Table 5 reports corresponding figures for external data whose sample is around 6 times the sample size for the internal data. BL3 accounts for 60.63% of the external data, BL4 for 6.46% and BL8 for 28.71%.

External data

	BL3	BL4	BL8
ET1	2.36%	0.04%	0.95%
ET2	62.11%	2.72%	0.17%
ET3	7.27%	0.34%	0.17%
ET4	6.27%	1.63%	40.81%
ET5	5.30%	0.10%	0.00%
ET6	0.74%	0.37%	0.28%
ET7	15.94%	5.46%	4.98%
	60.63%	6.46%	28.71%

[4] The seven event types are as follows. ET1: Internal Fraud, ET2: External Fraud, ET3: Employment Practices and Workplace Safety, ET4: Clients, Products, and Business Practices, ET5: Damage to Physical Assets, ET6: Business Disruption and System Failures, and ET7: Execution, Delivery, and Process Management.

Table 5: Description of External Database by business lines (BL3,BL4, and BL8) and event types (ET1 to ET7)

For the analysis of the data belonging to the three business lines and all events types, the internal data are sufficient as far as EVT is not concerned. This is not true, however, for the business lines that are not part of this study. Nevertheless, it is important to analyse the external data that are a benchmark for the bank and represent the operational risk of the entire financial system. Moreover, when investigating the behaviour in the tail, we need more data given that external data becomes essential when internal data sample is limited.

Mixing internal and external data must be performed with caution. Very often external data are influenced by predominant high events in the industry. Mixing internal and external data together may produce spurious results tending to be over-pessimistic regarding the actual severity distribution and leading to over-stated capital provisions for individual banks.

Our approach involves estimation of the distributions separately for internal and external losses that are collected with different thresholds, and then rescale the external data in order to make comparisons and aggregation meaningful.

The methodology applies to any loss sample, by business line, and relies on the following steps:

- Select a given distribution and estimate by the method of Maximum Likelihood the conditional distribution parameters of internal and external data separately,
- Standardize internal and external data using the the dispersion and location parameters from the fitted conditional distribution,
- Rescale the external losses by internal dispersion and location parameters,
- Compute the new threshold H as the rescaled value of the *max(internal threshold, external threshold*) and cancel out the data below the new threshold.

	External data	log(data)		Internal data	log(data)		Aggregated data	log(data)
BL3								
min	5000.00	8.52		500.00	6.21		1230.40	7.12
max	8547484.00	15.96		2965535.00	14.90		2965535.00	14.90
mean	37917.35	9.71		15888.12	8.12		11021.80	8.36
median	13200.00	9.49		2500.00	7.82		3406.98	8.13
std.dev	184787.23	0.99		97668.29	1.39		59968.38	1.03
skewness	26.53	1.29		20.17	1.02		26.60	1.32
kurtosis	910.88	5.23		516.07	4.16		953.26	5.41
BL4								
min	5000.00	8.52		500.00	6.21		521.54	6.26
max	20000000.00	16.81		1206330.00	14.00		2086142.29	14.55
mean	61880.06	9.73		28682.11	8.11		10175.51	7.59
median	12080.00	9.40		2235.60	7.71		1347.44	7.21
std.dev	653409.10	1.11		133873.92	1.67		83252.98	1.24
skewness	29.65	1.44		7.60	1.13		19.11	1.53
kurtosis	904.11	5.70		63.14	4.01		413.87	6.02
BL8								
min	5000.00	8.52		500.00	6.21		2396.93	7.78
max	15437000.00	16.55		606551.63	13.32		7400277.16	15.82
mean	40808.30	9.60		13089.05	8.11		19674.84	8.87
median	11000.00	9.31		2755.90	7.92		5273.24	8.57
std.dev	305524.49	0.99		48801.21	1.40		144912.26	0.99
skewness	35.67	1.59		9.49	0.89		35.87	1.58
kurtosis	1642.25	6.81		108.83	3.62		1668.56	6.71

Table 6. Descriptive statistics for internal and external operational loss data: BL3, BL4, and BL8.

From a statistical viewpoint, the integration of internal and external data is beneficial in particular for BL4 and BL8, where the internal data sample is limited. BL3 would also benefit in terms of statistical robustness by the database expansion, whereas internal data may be considered sufficient.

Without doubt, the external loss data influence the aggregate loss estimation for the three business lines: the mean, maximum and minimum losses, the skewness and kurtosis coefficients are higher for the external than internal data.

3.2. Loss severity and frequency analisys for the three business lines

In this section, we compare different parametric distributions fitted to the data from different sources employing the previously described procedure and focusing on the conditional estimates. We assume that the relevant frequency distribution is the one estimated on internal data. For the loss severity distribution, we will analyse separately internal and external data in order to find the distribution which better fits the empirical loss data. We use four criteria to select the best distribution:

1) graphical inspection of the cumulative distribution function, especially in the upper tail,
2) goodness-of-fit tests that we omit for brevity,
3) the convergence of the MLE procedure (value of the exit-flag =1),
4) the value *F(H)*, a proxy for the fraction of missing data (losses below H), must be acceptable for the bank.

After selection of the best distribution for internal and external data, we compute and analyze aggregate losses. For the aggregated data we present results for the Lognormal, generalised Pareto, Weibull, log-Weibull, and the Symmetric Stable distributions.

3.2.1 BL3 – SEVERITY AND FREQUENCY ANALYSIS

We include the results of the conditional severity distributions estimated separately on internal, external, and pooled data. Figures 1 and 2 present the graphical fittings on the tails of the conditional distributions estimated for the internal and external datasets. It is worth noting that the ranges for probability support are rather different for internal against external datasets: for the latter the x-axis goes up to $2.5 \cdot 10^6$, for the former to $14 \cdot 10^5$. Table 7 reports the ML estimates for the different distributions. The last column summarizes the ML coefficients for the aggregated data, also referred as total dataset (see Figure 3, here next). Missing entries in Table 7 indicate that the MLE procedure did not converge (exit flag=0). After applying our four criteria selection, the Symmetric Stable distribution provided the best fitting.

Following the methodology described earlier, we perform required standardization and rescaling of the external data. We then fit various distributions to the combined data.
We limit the analysis to the five selected distributions.

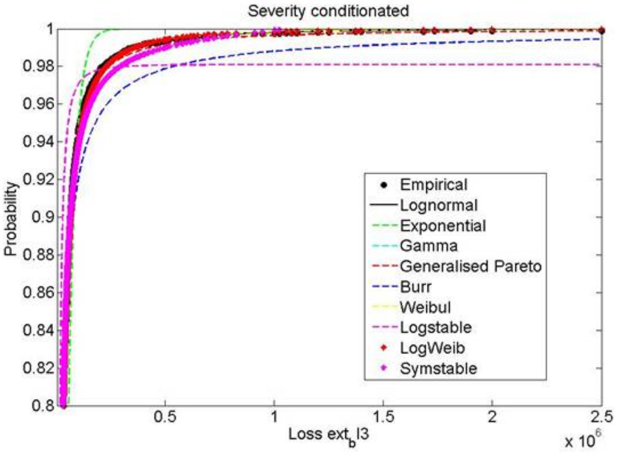


Figure 1. BL3 External loss distribution (upper tail)

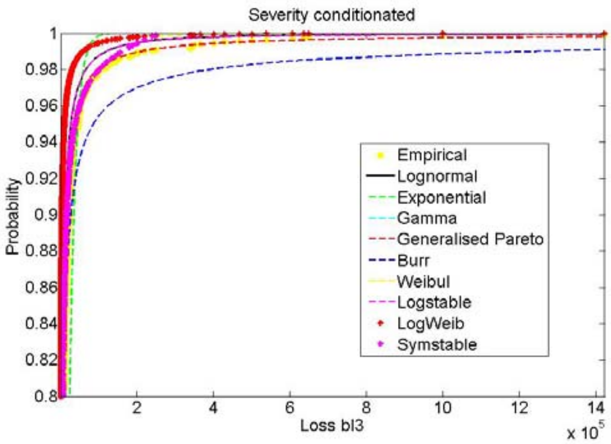


Figure 2. BL3 Internal loss distribution

	extern	intern	total
LN			
par1	7.15	6.69	5.54
par2	2.01	2.15	2.14
LL	-97985.62	-19262	-99912.5
F	0.75	0.41	0.77
exitflag	1.00	1.00	1.00
GPD			
par1	-0.80	-1.11	-0.86
par2	4.653.67	1.417.25	1.156.36
LL	-97994.46	-19264.5	-99919.6
F	0.54	0.26	0.53
exitflag	1.00	1.00	1.00
WEIBULL			
par1	0.53	0.31	0.85
par2	0.21	0.24	0.19
LL	-97994.01	-19270.5	-99923.2
F	0.96	0.76	0.96
exitflag	1.00	1.00	1.00
LogWEIBULL			
par1		0.00	0.01
par2		3.17	2.76
LL		-3141.89	-12528
F		0.47	0.80
exitflag		1.00	1.00
SymSTABLE			
par1	1.03	0.81	0.99
par2	6.410.28	1.577.44	1.597.92
LL	-20159.82	6567.48	17605.4
F	0.42	0.21	0.42
exitflag	1.00	1.00	1.00

Table 7. BL3 cond.l parameter estimation

Figure 3 demonstrates that collected losses are concentrated within the lower 90% of the probability distribution and few extreme losses are located in the upper 1% of the tail: the resulting parametric distribution are very steep up to the 95% and are almost indistinguishable, and then far in the tail the GPD and the SymStable appear to provide the best fit.

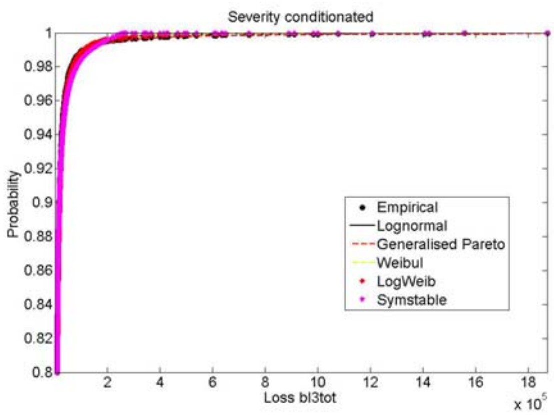


Figure 3. BL3 Total loss distribution (upper tail)

The frequency distribution is chosen taking into account internal data only: we again consider three assumptions[5] for the loss cumulative number of losses. The model providing the lowest approximation error, measured by the mean squared error (MSE), is chosen. Step size is chosen to be one week. Table 8 presents the parameter estimates for λ(t). For the Poisson case, par1 refers to λ.

	par1	par2	par3	par4	mse	mae
Type 1	97.7266	7846.17	1.15988	6.14317	1993.15	38.5637
Type II	2815.52	2717.47	1.2E-05	7.01935	1952.41	38.1942
Poisson	10.5079				20642.2	117.528

Table 8. BL3 Frequency distribution, internal data only

Figure 4 supports the results tabulated in Table 8 that homogenous or time-independent counting process is heavily inconsistent with the observed statistical evidence.

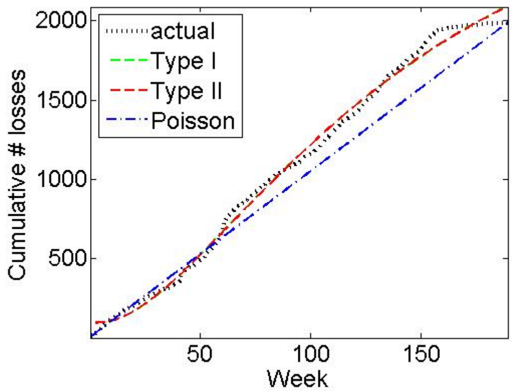


Figure 4. BL3 Loss events frequency distribution

The operational loss generating process, from the above evidence, is thus assumed to follow a non-homogenous Poisson process of Type II and a SymStable severity distribution. We select Symstable because the value of *F(H)* is more realistic.

3.2.2 BL4 – SEVERITY AND FREQUENCY ANALYSIS

A similar analysis can now be performed for the other two business lines. Consider first the operational losses generated by the Commercial Banking business line in which less then 200 losses have been recorded internally over a four year period with an impact above 500 euros each, while the external dataset includes less then 1,000 losses with a threshold of 5,000 euros. Determination of the best fit on the joint dataset again requires a standardisation procedure.

[5] Constant intensity with Poisson distribution, time dependent intensity with lognormal like (type I) frequency distribution, and time dependent intensity with logWeibull-like (type II) frequency distribution.

Figures 5 and 6 illustrate the fittings of the complete set of parametric distributions analysed on external and internal data respectively and Table 9 reports the associated ML estimation results for the considered distributions.

Following the information from Table 6, internal operational losses recorded during the sample period present high dispersion and asymmetry with respect to the other business lines, with rather extreme outliers. Internal loss data collection results in a rather irregular empirical distribution. The fitting on internal data, see Figure 6, is very bad, while the large sample of external data allows a good estimation procedure, see Figure 5. Missing entries in Table 9 indicate that the MLE procedure did not converge (exit flag=0) and the graphical evidence supports the lack of convergence specifically on internal data by the SymStable estimator.
The GPD proves to provide the best fitting on the internal and external dataset supporting results provided in Figure 7. However, the SymStable is the more appropriate distribution when the aggregated data are analysed.

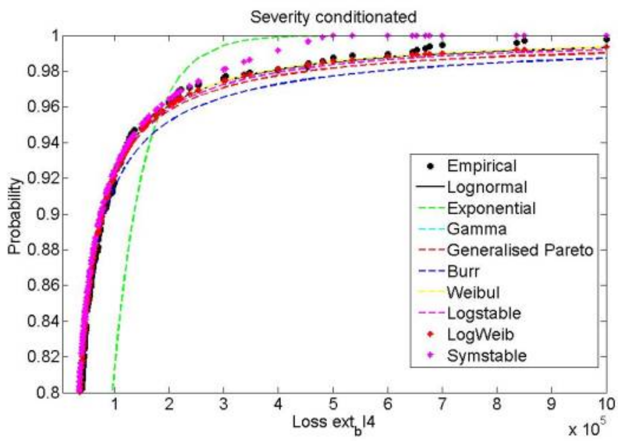


Figure 5. BL4 External loss distribution (upper tail)

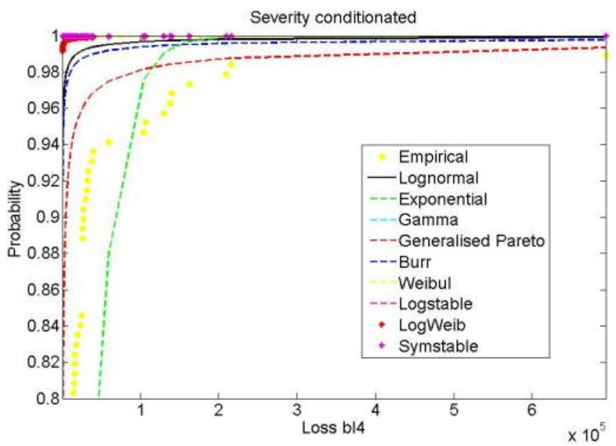


Figure 6. BL4 Internal loss distribution (upper tail)

	extern	intern	total
LN			
par1	0,8237006	-0,1582	-6,1183
par2	3,4751692	4,29817	4,45116
LL	-10499,28	-1831,5	-10108
F	0,9865803	0,93092	0,99728
exitflag	1,00	1,00	1,00
GPD			
par1	-1,097124	-1,8016	-1,2514
par2	1355,8954	141,43	100,661
LL	-10501,7	-1833,1	-10110
F	0,7712815	0,67	0,79979
exitflag	1,00	1,00	1,00
WEIBULL			
par1	3,8132002	2,6396	7,58354
par2	0,0900969	0,09408	0,06198
LL	-10499,09	-1831,3	-10107
F	0,9997292	0,99123	0,99999
exitflag	1,00	1,00	1,00
LogWEIBULL			
par1		0,06552	0,18929
par2		1,74192	1,48766
LL		-306,75	-1463,3
F		0,79386	0,94477
exitflag		1,00	1,00
SymSTABLE			
par1	0,8992893		0,83172
par2	3012,113		308,598
LL	-1512,135		-352134
F	0,6364079		0,62732
exitflag	1,00		1,00

Table 9. BL4 conditional parameter

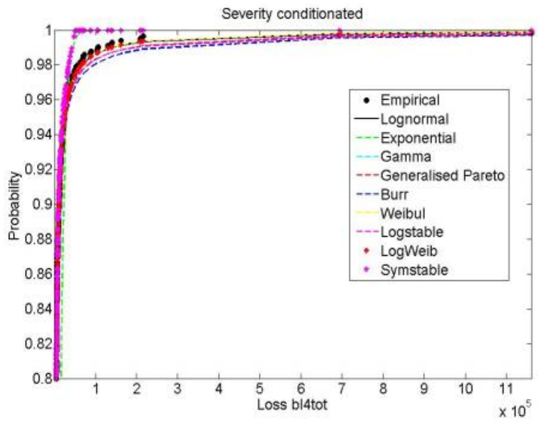


Figure 7. BL4 total loss distribution (upper tail)

Finally, we estimate the pooled data frequency distribution. Table 10 and Figure 8 support the choice of the non-homogeneous Poisson process of Type I for this BL.

	par1	par2	par3	par4	mse	mae
Type I	4.82938	1.4E+07	3.2591	18.9404	20.2075	3.50501
Type II	78928.3	78919.8	7E-08	6.44017	21.7972	3.74619
Poisson	1.24528				161.273	11.0271

Table 10. BL4 Frequency distribution, internal data only

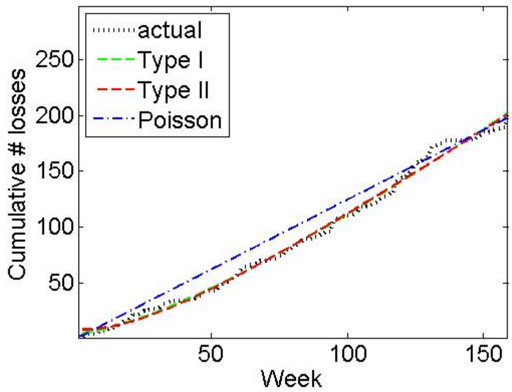


Figure 8. BL4 Loss events frequency distribution

We can conclude that for the losses recorded on this business line (all event types), although the GDP provides the best fitting on internal and external data, the SymStable is the distribution that fits the pooled data best. The preferable frequency distribution is the type I non-homogeneous.

3.2.3. BL8 – – SEVERITY AND FREQUENCY ANALYSIS

Finally, we analyse the collected evidence for the Retail Brokerage business line. As in previous sections, we first examine internal and external loss datasets independently and then pool the data across all event types. The dataset includes around 200 loss events recorded internally and around 4,000 events collected in the external database. All remarks on the statistical properties of internal versus external losses remain valid here. We consider first the two datasets separately.

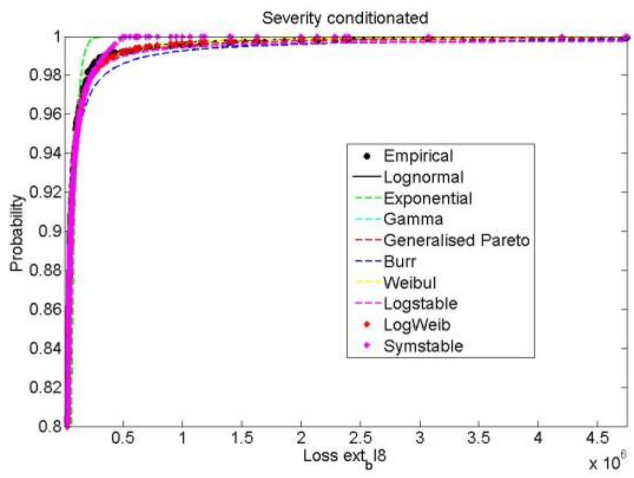


Figure 9. BL8 external loss distribution (upper tail)

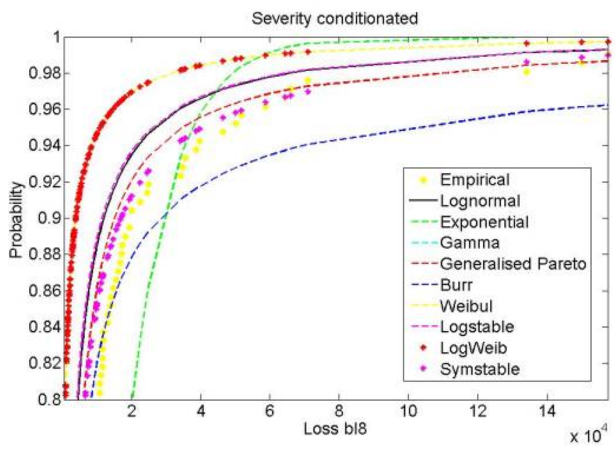


Figure 10. BL8 internal loss distribution (upper tail)

	extern	intern	total
LN			
par1	1.32515	6.57203	0.84834
par2	3.14546	2.20299	3.11317
LL	-43402	-2025	-41561
F	0.98889	0.43556	0.98703
exitflag	1.00	1.00	1.00
GPD			
par1	-0.9374	-1.1416	-0.9394
par2	1501.76	1340.21	750.878
LL	-43407	-2026.7	-41567
F	0.77924	0.26714	0.7713
exitflag	1.00	1.00	1.00
WEIBULL			
par1	4.1375	0.28588	4.20184
par2	0.09258	0.25253	0.0949
LL	-43402	-2025.1	-41562
F	0.99989	0.74673	0.99985
exitflag	1.00	1.00	1.00
LogWEIBULL			
par1		0.00198	
par2		3.16089	
LL		-330.24	
F		0.47183	
exitflag		1.00	
SymSTABLE			
par1	0.99234	0.78314	0.98861
par2	3067.49	1470.51	1498.07
LL	-7035.2	727.741	1241.82
F	0.6484	0.2284	0.64252
exitflag	1.00	1.00	1.00

Table 11. BL8 cond.l parameter estimation

Figures 9 and 10 illustrate the fit of a variety of loss distributions fitted, respectively, to external and internal datasets. In the first case again we have a rather large loss sample within the core of the distribution and then very large losses in the extremes: all distributions provide a good fit far from the tail and then on the tail, with the exception of the exponential, the others provide a good fit. On the contrary, for the internal losses only the SymStable distribution presents a sufficient fitting accuracy.

Goodness-of-fit tests suggest that none of the distributions provide a good fitting, but the estimates of the percentage of missing data (Table 11) support the choice of the SymStable for this business line.

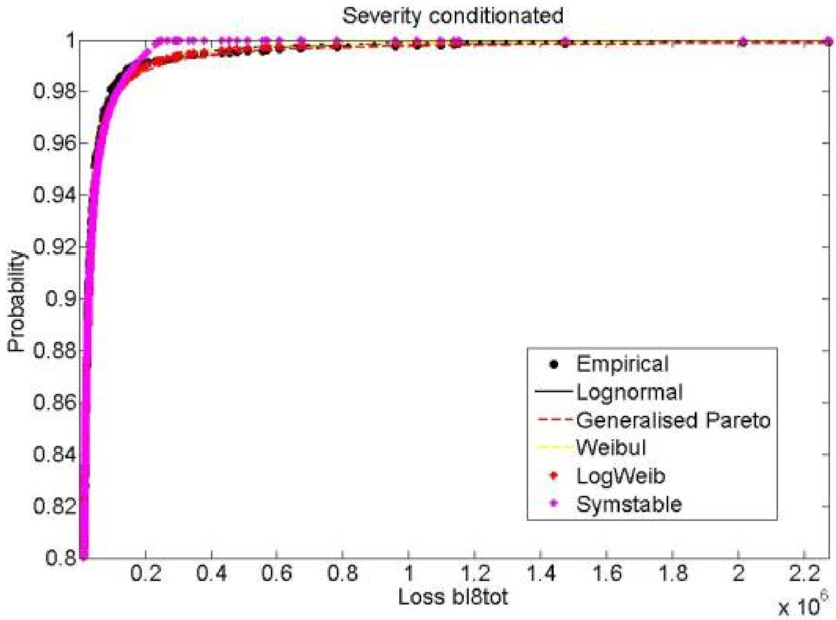


Figure 11. BL8 total loss distribution (upper tail)

Table 12 and Figure 12 report the numerical results and graphical evidence regarding the choice of the frequency distribution. The type II, logWeibull like, distribution is slightly more precise than the type I distribution as seen from lower error estimates as presented in Table 12.

	par1	par2	par3	par4	mse	mae
Type I	16.5887	695.9	1.17536	5.92135	20.4766	3.44717
Type II	258.197	241.009	3.2E-05	6.58742	19.877	3.35849
Poisson	1.14835				392.515	17.5465

Table 12. BL8 Frequency distribution, internal data only

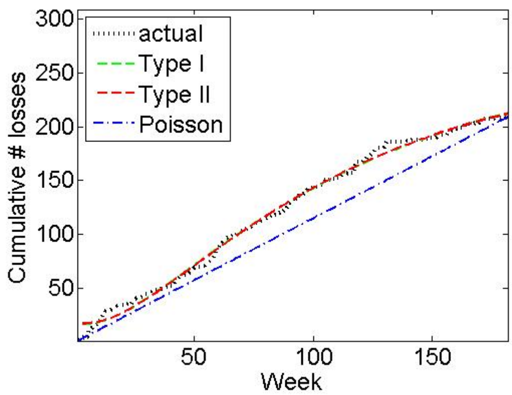


Figure 12. BL8 Loss events frequency distribution

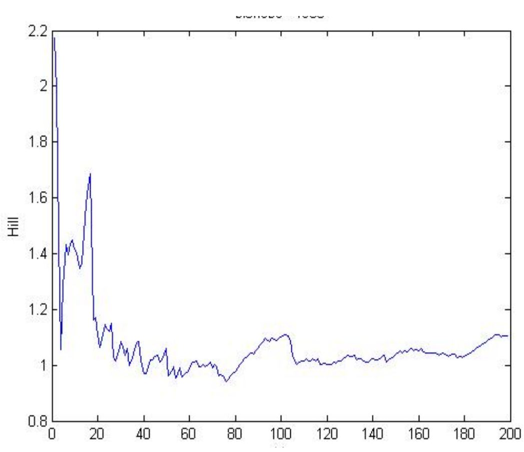
Figure 12 shows that for BL8 loss events are irregularly distributed across the different weeks, with from the best fit achieved when the non-homogenous model is used instead of the time homogeneous model. Summarizing, the analysis legitimizes the selection of the SymStable distribution as preferable severity distribution and the type II frequency model.

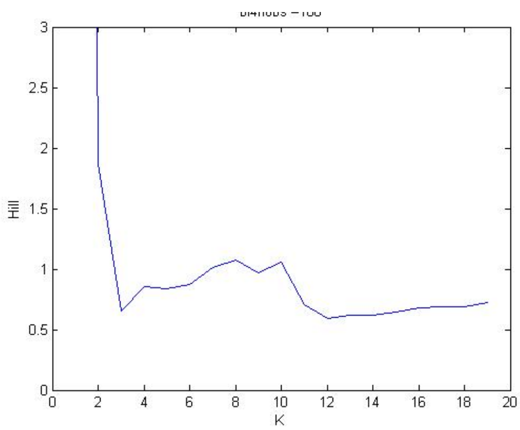
3.3 MIXED SEVERITY DISTRIBUTION

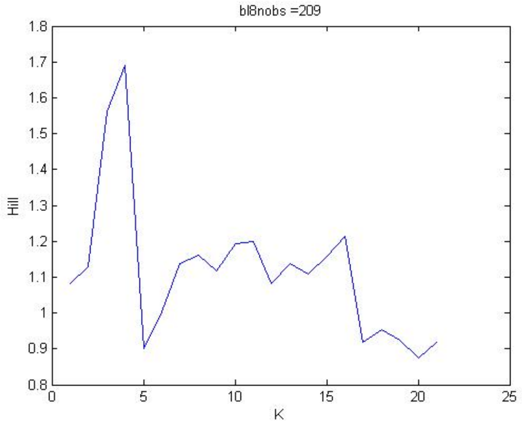
As a final step, we investigate the possibility to model the tail of the severity distribution by EVT.
The idea is to model the body of the distribution for each business line with the best distribution that we have identified in the previous step and the tail with a GPD. As mentioned in Section 2, the problem is solved introducing a mixed distribution.
Let *u* be the selected threshold for the support partition that identify the beginning of the tail, we then assume:

$$h(x) = \pi \frac{f(x)}{F(x)} 1_{x \le u} + (1-\pi) g(x) 1_{x > u}$$

where in our applications, following the results in Section 3.2, *f(x)* is the body distribution and *g*(x) is the GPD. We note that the two distributions work on disjoint datasets once the threshold is identified. Several techniques can be employed to isolate such critical value: all of them, however, require a sufficient dataset of extreme losses and implies thus the solution of a trade-off problem. The extreme loss events exceeding *u* for BL3 in the internal dataset are sufficient to perform with the required statistical accuracy the GPD ML estimation procedure, while in the case of BL4 and BL8 both internal and external data must be used for the solution of the identification problem. In Figure 13 we report the evidence on the threshold selection on the internal data only for BL3, BL4 and BL8.







Top left: the estimation iterative procedure for BL3 for increasing threshold does converge to a sufficiently stable value with a sample of roughly 50 losses: the convergence is studied through the Hill estimator, common in this type of applications. On the contrary, top right and bottom, respectively for BL4 and BL8, we do not have any convergence of the estimator on the limited internal dataset suggesting the adoption of the integrated dataset.

Figure 13. Selection of the extreme threshold for the GPD fitting far in the tail

Following the indications in Section 3, we will rely on the total or pooled dataset and again consider the two options for all business lines: the single SymStable or the mixed SymStable-GPD as parametric severity distribution.

Given a threshold *u*, it is possible to estimate π using ML in order to identify the optimal convex combination of the SymStyable density *f(x)* and the GPD *g(x)*. Table 13 defines the outcome of this procedure and provides an input to the Monte Carlo scenario generator. In Table 13 we report initial points *F(u)* of the optimization procedure; however, we tried different starting points. The optimal solutions turn out to be robust.

	-LogL	π
BL3		
Starting point	-1.6765	0.9922
Optimal	-1.677	0.9901
BL4		
Starting point	3.378	0.9871
Optimal	3.377	0.957
BL8		
Starting point	0.0614	0.9799
Optimal	0.0552	0.9512

Table 13. Optimal partition of the loss events for the mixed distribution

Relying on the above results we can now consider both procedures for the generation of the marginal distributions, to be interfaced with the copula estimator for the estimation at a given risk horizon of the global operational loss distribution for the bank. On the right column of Table 13, we also have now a relative measure of the number of scenarios, relative to the total simulated loss scenarios, that generate the very extreme losses, for which the GPD severity distribution must be sampled.

4. CONCLUSIONS

In this paper we discuss an approaches suggested by Basel II to compute the operational risk capital charge. In the presence of a reasonable number of data (i.e., losses classified by

business line and event type), the only approach that properly considers the characteristics of the bank data is the LDA approach.
The LDA methodology allows to estimate the frequency and severity distributions separately for each "business line/event type" combination.
For the severity distribution, we use a large number of candidate distributions and propose a methodology to select the one which fits the observed data best. At the same time the frequency model is selected among three different models with constant intensity or deterministically varying intensity.

For the three business lines used in the empirical study, that account for large part of the bank's operational losses, the following severity distributions and frequency models have been determined as the preferable marginal distributions.

	External	*Internal*	*Pooled*	**Frequency**
BL3	SymStable	SymStable	SymStable	Type II
BL4	GPD	GPD	SymStable	Type I
BL8	SymStable	SymStable	SymStable	Type II

Table 14. Final evidence on the BL's preferable parametric approximations

Overall, the SymStable distribution provides the best parametric fitting on all three business lines once internal and external data are pooled. With respect to the frequency distribution, the most accurate frequency model for all three business lines, the non-homogeneous frequency model, a result that agrees with findings in our previous studies.

Once that the best combination severity/intensity has being selected, we can compute via Monte Carlo simulation the distribution of the aggregated losses and the capital requirement as the VaR and the Cvar of the distribution at the 99.9% confidence level for each BL separately. Namely we can generate the marginal distribution of the losses deriving from different BL. Basel II suggests to add up the individual VaR figures, however the inclusion of a copula function will permit to move from an underlying perfectly correlated assumption to a system in which the possibility of more realistic correlated operational events is taken into account.

REFERENCES

Balkema, A. and de Haan, L. (1974), "Residual life time at great age", *Annual of Probability*, **2**, 792-804.

Baud, N., Frachot, A. and Roncalli, T., 2002, "Internal data, external data and consortium data for operational risk measurement: how to poll data properly", *Technical Report*, Crédit Lyonnais, Groupe de Recherche Opérationelle.

Bee, M., 2005, "On maximum likelihood estimation of operational loss distributions", *Technical Report*, **3**, University of Trento.

Bening, V. E., and V. Y. Korolev, 2002, *Generalized Poisson Models and their Applications in Insurance and Finance* (Utrecht, Boston: VSP International Science Publishers).

BIS, 1998, Operational Risk Management, www.bis.org.

BIS, 1998, Overview to the Amendment to the Capital Accord to Incorporate Market Risks, www.bis.org.

BIS, 2001a, Consultative Document: Operational Risk, www.bis.org.

BIS, 2001b, Working Paper on the Regulatory Treatment of Operational Risk, www.bis.org.

BIS, 2003, "The 2002 Loss Data Collection Exercise for Operational Risk: Summary of the Data Collected", www.bis.org.

BIS, 2004, "International convergence of capital measurement and capital standards", www.bis.org.

Chernobai, A., Burnecki, K., Rachev, S., Trück, S. and Weron. R., 2006, "Modelling catastrophe claims with left-truncated severity distributions", *Computational Statistics*, 21(3).

Chernobai, A., Menn, C., Trück, S. and Rachev, S., 2005b, "A note on the estimation of the frequency and severity distribution of operational losses", *Mathematical Scientist*, **30**(2).

Chernobai, A., Rachev, S., Trück, S. and Menn, C. , 2005c, "Estimation of operational value-at-risk in the presence of minimum collection thresholds", *Technical report*, University of California, Santa Barbara.

Chernobai, A., Rachev, S. and Fabozzi, F.,2005d, "Composite goodness-of-fit tests for left truncated loss samples", *Technical report*, University of California Santa Barbara.

Chernobai, A., and Rachev, S., 2006, "Applying robust methods to operational risk modeling", *The Journal of Operational Risk*, **1**(1).

Crouhy, M., D. Galai, and R. Mark, 2001, *Risk management* (New York: McGraw-Hill).

Cruz, M. G., 2002, *Modeling, Measuring and Hedging Operational Risk* (Chichester, New York: John Wiley & Sons).

Dempster, A. P. , Laird, N. M. and Rubin, D. B., 1977, "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society*, Series B (Methodological), **39**(1), 1–38.

Embrechts, P., C. Klüppelberg and T. Mikosch, 1997, *Modeling Extremal Events for Insurance and Finance* (Berlin: Springer-Verlag).

Grandell, J., 1991, *Aspects of Risk Theory* (New York: Springer-Verlag).

Grandell, J., 1997, *Mixed Poisson Processes* (London: Chapman & Hall).

Huber, P.J., 2004, *Robust Statistics*, (Hoboken:John Wiley & Sons).

King, J.L., 2001, *Operational Risk: Measurement and Modelling* (New York: John Wiley & Sons).

Knez, P.J. and Ready, M. J., 1997, "On the robustness of size and book-to-market in cross-sectional regressions", *Journal of Finance*, **52,**1355–1382.

Jorion, P., 2000, *Value-at-Risk: The New Benchmark for Managing Financial Risk,* Second

Edition (New York: McGraw-Hill).
Martin, R. D. and Simin, T. T., 2003, "Outlier resistant estimates of beta", *Financial Analysts Journal*, **59,** 56–69.
McLachlan, G. and Krishnan, T., 1997, "The EM Algorithm and Extensions", *Wiley Series in Probability and Statistics*, (John Wiley & Sons).
Meng, X.L. and van Dyk, D., 1997, "The em algorithm - an old folk-song sung to a fast new tune", *Journal of the Royal Statistical Society*, Series B (Methodological), **59**(3), 511–567.
Moscadelli, M., Chernobai, A.and Rachev S. (2005) Treatment of incomplete data: the effects on parameter estimate, EL and Ul figures, www.operationalriskonline.com, June.
Pickands, J. (1975), Statistical inference using extreme order statistics, *Annals of Statistics*, 3, 119-131.
Panjer, H. H. . and Willmot, G., 1992, Insurance Risk Models. Society of Actuaries, Schaumburg, Illinois.
Peemöller, F.A., (2002), Operational risk data pooling, Deutsche Bank AG, Presentation at CSFforum- Operational Risk, Frankfurt/Main.
Rachev, S. T. (ed), 2003, *Handbook of Heavy Tailed Distributions in Finance. Book 1*. In North Holland handbooks of finance (Series editor W. T. Ziemba) (Amsterdam, Boston, London, New York: Elsevier).
Rachev, S. T., and S. Mittnik, 2000, *Stable Paretian Models in finance* (New York: John Wiley & Sons).
Rachev, S. T., S. Mittnik, and M. Paolella, 1998, "Stable Paretian Modeling in Finance: Some Empirical and Theoretical Aspects", in R. J. Adler, R. E. Feldman, and M. S. Taqqu (eds), *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, pp. 79–110 (Boston: Birkhäuser).
Resnick, S. I., 1987, *Extreme Values, Regular Variation, and Point Processes* (New York: Springer-Verlag).
Roehr, A., 2002, "Modelling operational losses", *Algo Research Quarterly*, **5**(2), 53-64.
Rousseeuw, P. J. and Leroy, A. M., 2003, "Robust Regression and Outlier Detection" (Hoboken: John Wiley & Sons).
Samorodnitsky, G., and M. S. Taqqu, 1994, *Stable non-Gaussian Random Processes. Stochastic Models with Infinite Variance* (London: Chapman & Hall).
Thorin, O., and N. Wikstad, 1977, Calculation of Ruin Probabilities when the Claim Distribution is Lognormal", *Astin Bulletin*, **9,** pp. 231–46.
Willmot, G., 1990, "Asymptotic Tail Behaviour of Poisson Mixtures with Applications", *Advances in Applied Probability*, **22,** pp. 147–59.